LIBRARY
TECHNIQUE
NAVAL
MONTEREY CALIFORNIA 93

NPS-55Mt73121

NAVAL POSTGRADUATE SCHOOL

Monterey, California

A MANPOWER AND PERSONNEL RESEARCH

DATA BASE FEASIBILITY STUDY

Kneale T. Marshall

December 1973

Final Report for Period January 1973-December 1973

Approved for public release; distribution unlimited.

Prepared for:
Office of Naval Research, Arlington, Virginia 22217

FEDDOCS
D 208.14/2:
NPS-55MT73121

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral Mason Freeman
Superintendent

Jack R. Borsting
Provost

The work reported herein was supported by the Office of Naval Research.

Reproduction of all or part of this report is authorized.

This report was prepared by:

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: NPS55MT73121A NPS-55Mt73121	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE *(and Subtitle)*: A MANPOWER AND PERSONNEL RESEARCH DATA BASE FEASIBILITY STUDY	5. TYPE OF REPORT & PERIOD COVERED: Final Report Jan 73-Dec 73
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): K. T. Marshall	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Postgraduate School, Monterey, CA 93940	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Project No 3-0131
11. CONTROLLING OFFICE NAME AND ADDRESS: Office of Naval Research	12. REPORT DATE: Dec 1973
	13. NUMBER OF PAGES: 45
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*	15. SECURITY CLASS. *(of this report)*: Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for Public Release; distribution unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Manpower Research Data Base
Data Base
Personnel.

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

Manpower and Personnel Data bases are classified into operating, staff and research types. Characteristics of each type are discussed. The demand for a research data base is discussed and approximate costs of alternatives are given. Data base structure, implementation and development are each considered together with the location of a proposed research data base.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

ACKNOWLEDGEMENT

The author would like to thank Dr. R. A. Weitzman for his help in the development of questionnaires, and in their administration, collection, and collation.

I. INTRODUCTION

Manpower and personnel research has commanded increased interest and support in recent years as a result of a number of factors. Great impetus was given by the advent of the "All Volunteer Force" (AVF) concept to replace the draft, and by the rapid increase in manpower costs which resulted from it. Research into testing and training programs has been supported for some time. This is now being expanded, new areas of attitudinal research are being investigated, and more emphasis is being placed on force planning models. The Ginsberg Report [1] can be cited as one example where an increase in research support for manpower was recommended.

As a result of the increase in interest in these areas in the Navy in general, and as a result of the research sponsored by the Office of Naval Research in particular, it became clear that many researchers would be looking for manpower and personnel data to support their work. The total Navy annual cost of manpower and directly related areas is about $12 billion. The total annual expenditure on manpower and personnel research (see Table 1) is approximately $59 million, or approximately ½% of total manpower costs. Although small in comparison to total manpower costs it is a significant amount, and such research activity both requires and generates important amounts of data. One need only look at the number of surveys made in the last few years in personnel areas to realize the types and quantities of data generated by research and development people. There is currently no central location for storing and making available the results of such surveys on Navy-related problems. The concept thus arose of a centralized data base for manpower and

personnel research, a data base which would act as a repository for historic data and which could be accessed with a minimum of difficulty by both civilian and military research institutions. Such an idea is of course not new, and in the department of defense, for example, there is currently in existence the Manpower Research and Data Analysis Center (MARDAC) which maintains very extensive historic data files from a number of sources. The purpose of this report is to clarify certain concepts related to data bases and to present various alternative ways in which a Navy Manpower Research Data Base might be operated. In section II data bases are partitioned into three types and the characteristics of each type are discussed with examples. In Section III the question of demand for a research data base is discussed. In section IV the problems of costs and some specific possible data base systems are considered. In section V some remarks regarding the structure of a research data base are presented, and in section VI the implementation and development of a research data base are discussed. Two possible locations are considered in VII and in VIII some concluding remarks are made.

	Budget ($1,000,000)
Human Resources	26.8
Biological and Medical Sciences	28.6
Laboratory Supplies and Instrumentation	3.5
TOTAL	58.9

Table 1. Manpower, Personnel, Training and Medical Research Budget for FY 1974*.

(Source: R&D Plans Division, Office of Research, Development, Test and Evaluation, Office of the Chief of Naval Operations).

II. DATA BASE TYPES AND CHARACTERISTICS

In this report we consider data bases as being in one of three categories. These are named 1) Operating, 2) Staff and 3) Research and these terms are explained in some detail in this section. The reader is warned that the word "research" is used in a very specific sense in this report to refer to work of a fundamental nature that involves new theory and increases the understanding of manpower and personnel problems. This usage is much narrower than the usual one where it is used to describe much activity which is not considered directly operational. By narrowing the use of the word it is easier to discuss the similarities and differences of the three data base types.

The three types of data base are described and categorized by certain characteristics which are either inherent in their make-up or are desirable as properties. These descriptions help to clarify the rather large differences in the various data base types.

1. Dynamic versus Static Data.

Operating data bases (ODB's) exist to aid the personnel planners and managers in the recruitment, training, assignment, movement, promotion and payment of personnel. In a force of 500,000 men each of which is reassigned jobs on the average of every two years, approximately 700 people change jobs each day. If the average time spent in the Navy is about 5 years the daily recruitment rate to maintain the force at the half million level is almost 300. Numerous other personnel changes occur daily also, such as assignment to schools or training centers, completion of training, promotions, changes in dependant status etc.

All such information on both officers and enlisted men is captured on a centralized data base in the Bureau of Naval Personnel (Pers N) and is stored on magnetic tape. The enlisted data comprises about 1.5 billion bytes of information and each day about 10,000 changes are made to personnel records on the tape. It is obvious that this operating data base is very dynamic and is in a constant state of change. This is an inherent characteristic of all operating data bases.

In this report a research data base (RDB) refers to a data base intended for use primarily by research organizations such as civilian universities and research institutions, or Navy research laboratories and institutions such as the Naval Postgraduate School. Studies undertaken by such institutions are usually of a long-term nature (1 or 2 years) and the data base to which they would require access would consist of static data recorded at some earlier time point. There would be no important requirement to update the data base constantly as in the case of the ODB as it would not be used for daily decision making.

The third type of data base considered is called a staff data base (SDB) which is in some sense in an intermediate position between the ODB and RDB. An example of an SDB is MARDAC (ref [2]). The overwhelming use of the SDB is in support of a staff of senior decision makers who require rapid answers in terms of statistical tabulations of historic performance. Such a data base is not dynamic in the same sense as an ODB. Although new data is constantly being added to the SDB on a regular basis, few if any changes are made of the type inherent in the ODB. Although far more additions of new data would be made in any time period to the SDB than the RDB one can consider them both as being static.

2. Current versus Historic Data.

An ODB is of value only if it contains current data. It must contain records of current enlisted personnel and maintain these records in as up-to-date form as possible. An RDB should have historic information which can be drawn upon to study the effects of past and, hopefully, future policies. It is not important that it contain the latest available information. Similarly an SDB need not be completely current, though its use calls for more recent data than would be available in an RDB. For example the SDB may get a request for data on the racial breakdown of recruits from certain geographical areas by month for the previous six months. Such data would not normally be available from the ODB since it contains only current data. The data might be available from an RDB but the RDB is not designed to answer such questions on a routine basis and may not be current enough.

3. Cross-sectional versus Longitudinal Data.

The ODB deals almost exclusively with cross-sectional data. It contains a"snapshot" of the Navy at some timepoint of who is currently in service, in what status, and where. It does contain limited historical information on each individual such as his date(s) of entry into service, his last duty assignment and a few more pieces of history which determine his qualifications. Of course, his entire service record is available on flat paper, but here we are discussing rapid access computer-based data bases. It would not be economically justifiable (nor desirable) to put an individual's entire record into a rapid access system. The researcher, in contrast to the operator, is often interested in behavioral characteristics of groups, or cohorts. An example might be the behavior of two

groups, one of which had taken a particular training course on drug abuse, and a control group which had not gone through the training program. Information would be needed on each group as it progressed in time. This is what we call longitudinal data and any RDB would typically have this kind of data.

The SDB usually contains both cross-sectional and longitudinal data. The ones with which the author is familiar collect cross-sectional data from ODB's at certain time points such as end-of-month or end-of-year, and create longitudinal data by merging and sorting of these snapshot records. Such operations, although simple in concept, require a heavy commitment of computer and personnel resources and typically cannot be carried out by an ODB.

4. Rapid versus Slow Access.

An ODB must possess rapid access. The current Pers N data base gives overnight service on most operating needs. However, a study is currently underway to determine the feasibility of an on-line data base which could be accessed by remote terminal, in which case access times would be in the order of minutes and seconds in place of the current hours and days (see ref. [3]). Such rapid access of all the data in an RDB would not be necessary. Most data could be stored off-line since the time pressures on researchers are in no way similar to those of operators. The reader should not interpret these remarks as saying that a researcher should not have access to an on-line system. Indeed he may do most of his work on such a system and any data he needs can be fed into it when appropriate. We are simply pointing out here that whereas an operator needs rapid access constantly to all the data, the researcher needs access to only

a small part of a research data base occasionally so that the RDB should be stored off-line in some kind of library format.

The SDB is again somewhere between these extremes. Most staff people demand rapid answers to their questions so that rapid access is desirable. This ability to rapidly answer questions may be the justification for the existence of the data base. However, the SDB is usually considerably larger than the ODB so that it would be impossible to justify (and probably technically not feasible to accomplish) the on-line storage of such data. The SDB typically requires rapid access in the order of hours or days by use of high speed tape units.

5. Population versus Samples.

The ODB must maintain information on every individual. One of its main uses is in the assignment of individuals to billets and thus records must be kept on the whole population.

The RDB is often used for work on inference. The effects of various policies on past behavior are often investigated and for these it may be sufficient and desirable to have available a sample of personnel records rather than those of the entire population.

The suitability of sample data for the SDB is not quite clear. It would appear that samples would be adequate for answering a large majority of the questions put to such a group. However, the questions at times require statistics on such small groups that, without data on the entire population, no one can be found in the required groups. It would appear that if the groups are in fact so small, then statistics on them would be of little interest. This does not seem to be the case, and this point helps to distinguish further the differences between a

SDB which exists primarily to answer the questions of high level staff, and an RDB which is intended to be of use to serious researchers.

6. Single System versus Multiple System.

An ODB typically operates in a single computer system to which operators have various kinds of access. They may request data in the form of reports issued by an input/output division or through remote terminals. The main point here is that the data remains in the same computer system almost exclusively.

An RDB may be accessed by many researchers from different institutions with different hardware and software configurations. In fact, it is not known in what system the data may be used in the future. Such uncertainties affect the way in which one should consider storing data. Both the SDB and the RDB have the problem of collecting data from various sources and merging and editing for their own system. However, because of the large internal data use by the SDB it typically should not concern itself with compatibility of other systems to the extent that the RDB must.

Table 2 summarises the comments in this section. The entries should be interpreted not as hard and fast requirements but simply as pointing out where different types of data bases have different emphasis.

To end this section we mention a characteristic desirable of all data bases. That is that they should all have EASY access. This may seem self-evident, but anyone who has experienced the frustrations of accessing tapes of data, and having done so must then gain confidence in and understanding of the contents, will appreciate the significance of these comments. There are many billions of bytes of information

currently stored on tape which probably will never be accessed. The reasons are numerous, but an important one is that access is so difficult that many researchers get frustrated and move on to other problems. Poor documentation, formating problems, and differences in computer operating systems are some of the main causes. If an RDB is to be of use it must overcome these problems.

TYPE OF DATA BASE

Characteristic	ODB	SDB	RDB
Dynamic Data	Yes	No	No
Static Data	No	Yes	Yes
Current Data	Required	Not Required	Not Required
Historic Data	Small Amount	All	All
Cross-section Data	All	Some	Some
Longitudinal Data	None	Some	Most
Rapid Access	Required	Sometimes	Not Required
Population Data	Required	?	Not Required
Sample Data	Not Sufficient	Sufficient	Sufficient
Single System	Yes	Yes	No
Easy Access	Yes	Yes	Yes

Table 2: Summary of Data Base Characteristics

III. THE DEMAND FOR A RESEARCH DATA BASE.

In the previous section characteristics of an RDB are discussed which distinguish it from other data bases. There is no doubt that an RDB in manpower and personnel would be of help in certain research projects, but as in all proposals to institute a new service, some attempt must be made to estimate the demand for such a service. Also, an attempt must be made to determine the types of data most often desired by researchers. In this section, we describe our efforts to answer these questions.

The fact that this study is being undertaken is evidence of some interest in an RDB. Thus we started our efforts in this area by assuming interest on the part of the research contractors funded by ONR under the AVF program and mailing out questionnaires to each contractor. The contractors were 1) Stanford Research Institute, 2) American Institutes for Research, 3) George Washington University 4) Institute of Social Research, University of Michigan, 5) University of California, Berkeley.

Two questionnaires were circulated, one requesting very detailed information on data elements and one which was restricted to a few (5) items (see Figure III-1) of a more general nature. The answers to the detailed questionnaire were neither complete nor revealing. The answers to the shorter questionnaire were much more complete and useful and are summarized here.

The answers to item 1 are listed in Figure III-2. These answers are not verbatim quotes, but have been modified for clarification. One can immediately see the large diversity in the research projects. Of

NAME_____

AFFILIATION_____

Questions on Data Requirements of Research on Personnel and Manpower Problems of the Navy

1. Briefly describe the objective(s) of your research (e.g. [1] modeling of force structure and change, [2] prediction of individual behavior, etc.).

2. For the data used in your research to date:

   a. provide a brief description of the data (e.g. [1] continuation fractions by rating, [2] date of enlistment, etc.),

   b. tell where you obtained the data

      (1) agency of the Navy (specify)
      (2) civilian agency (specify)
      (3) your own survey
      (4) other (specify),

   c. indicate if the data obtained adequately met your requirements,

   d. specify any problems that you had in obtaining the data.

3. Describe briefly the data that you would like but have not been able to get.

4. Considering your response to question 1, please indicate whether your data requirements consist primarily of

   a. descriptions of individuals or

   b. statistical descriptions of groups.

5. In the preceding questions, we may not have asked for complete information on your specific data requirements. Please list now any additional data requirements that you may have or anticipate having in the foreseeable future.

Figure III-1: Questionnaire mailed to AVF Researchers

1. To determine the impacts of alternative instructional technologies for Navy enlisted men both on costs and on the force itself.

2. To model the Navy enlisted man's career by a decision tree and evaluate the relative effectiveness of different personal and organizational incentives.

3. To analyze the effects various economic variables have on enlisted manpower supply in the Navy and Marine Corps in All Volunteer Force environment.

4. To project the organizational-change requirements of an All Volunteer Navy from value and work-preference changes in society.

5. To model movement of manpower in and between skill categories and to develop interactive computer models to aid manpower policy planners in determining the effects of various manpower policies on requirements and retraining requirements.

Figure III-2: Answers to Item 1: Briefly describe objectives of research.

course, many other research projects are currently under way by other groups and we did not attempt to cover all projects. Conversations were held with the Naval Personnel Research Lab in San Diego, with CNA, and with a number of other groups. The main conclusion to be drawn is that the research areas encompassed by the terms "manpower and personnel" are very broad but with heavy emphasis on behavioral and psychological areas.

The answers to item 2a show that the data requirements could be grouped into 3 categories. The five responses are given in Figure III-3, but these can be summarized by grouping the data into

1) Responses to researchers' questionnaires.
2) Historical and current statistics on the force structure and requirements.
3) Cost break-down data.

Category (1) data is clearly generated by the researcher himself, and duplication of effort can easily occur if different research groups who are working in similar areas are unaware of each other's survey efforts. There is some evidence that such duplication is quite common (see ref [4]). Although the results of surveys are often to be found in technical reports, it is also true that many times only a small fraction of the data collected on a survey finds its way into these reports. It may also be true that the data is no longer in a form useful to other researchers, who must then institute a further survey. This problem will be pursued in section VI. Clearly the answers to surveys form one important category of the researchers' data requirements.

Category (2) is probably the easiest type of data to obtain from

The operating data bases. Snapshot information is the easiest to obtain on force structure and requirements. One must manipulate data from various time points to obtain longitudinal data, but this is possible and is done to calculate, for example, continuation and loss rates.

Category (3) is in some ways similar to category (2) but is treated as a separate category since it is much more difficult to collect or determine this type of data. When a researcher asks the cost of training a man in a given skill at a given school he should understand the difficulties of answering such a question. Such data is not usually readily available and is often the output of mathematical models developed in other research programs, and is not straightforward historical accounting data.

The answers to item 2b show the diversity of sources from which the researchers obtained the data. Since much of the research work involved enlistment and reenlistment incentives, civilian data as well as Navy data was required. Figure III-4 summarizes the sources of data determined by the questionnaire.

The answers to item 2c show that in most cases the data met requirements. Some difficulties occurred using data generated in operating systems to match with survey results, but the answers pointed out no serious or unexpected problems.

Various problems experienced in obtaining data were found in answers to item 2d. These are summarized in Figure III-5. An important problem which arose was in the administering of survey questionnaires. Many groups are currently making surveys in the fleet and this may have led to the difficulties. We pursue this point later. Identification of

1. Costs of maintaining a man in each rating at each rate, costs of training him, overhead, retirement, etc.

2. Questionnaire responses to hypothetical situations.

3. Re-enlistment rates, structure of recruitment population, effects of variable re-enlistment bonuses, occupational structure of civilian and military populations, benefits (monetary and non-monetary) of military service, length of enlistments and promotion data.

4. Questionnaire responses indicating perceptions and preferences regarding practices and conditions of military service, and performance data on questionnaire respondents.

5. Current legacies and force structure by rating, past continuation rates, future requirements by rating.

Figure III-3: Answer to Item 2a: Give a brief description of data required.

1. Bureau of Naval Personnel (Pers Ax) for rating costs, and Memphis for training costs.

2. Survey questionnaires administered at AFEE stations, junior colleges and to first term enlistees.

3. Marine Corps Cohort File, Headquarters Marine Corps, Navy Enlisted Master Tape, Bureau of Naval Personnel (Pers N), Bureau of the Census, Marine Corps recruit survey data, Congressional hearings and DOD force structure data.

4. Performance data from Bupers, and survey questionnaires.

5. Requirements data from Op 01b, legacy data from Bupers (Pers N) and the Naval Personnel Training and Research Lab, San Diego.

Figure III-4: Answers to Item 2b: From where was the data obtained?

1. Some difficulty in identifying from where the data could be obtained, especially the cost data.

2. Lags in updating the master tape to problems of correlating master tape data with survey results. Concurrent studies involving surveys interfered with subjects of survey.

3. Delays in obtaining data affected the sequencing but not the results of the research.

4. Difficulty experienced in obtaining OMB approval for survey questionnaire. Problems using operating data due to methods of aggregation and storage.

5. Problems discerning the exact nature of the data and having confidence in its correctness, mostly due to documentation and communication problems.

Figure III-5: Answers to Item 2d: Indicate problems in obtaining data.

sources is a recurring problem and ref [5] is an attempt to reduce difficulties in this area.

The answers to item 3 should have a direct bearing on any decision to implement a research data base. One of the main objectives of such a decision would be to make available data which cannot be obtained readily from the operating system or the research laboratories. The answers are given in Figure III-6.

Although the answers to item 3 point out some problems, the reader is cautioned that such answers cannot give a complete picture of data requirements. Researchers are frequently restricted to working on problems for which data is available, and they often would like to move into interesting and relevant areas, if work in these areas would be fruitful. Lack of data for model validation or operation can make work in certain areas no more than of academic interest.

The answer to item 4 show that two of the research groups (1 and 5) did not require data on individuals while the others (2, 3 and 4) required data on both individuals and groups. There is obviously a need for both kinds of data.

The answers to item 5 were varied, but they indicated a need for longitudinal cohort-type data from the Navy similar to the Marine Corps cohort file (ref [6]). In a letter to Dr. R. A. Weitzman, Dr. A. S. Glickman, Deputy Director of the Washington Office of the American Institutes for Research, says: "...that we have found over the years that the principal deficiency is in data permitting longitudinal studies; i.e., individual data accumulated consistently over time in the Navy. Since, for the most part, as new entries are made to a persons record,

1. None.

2. Generated data, resulting, for example, from actual experiments in which enlistees are allowed to leave when they want to.

3. More detailed census data (e.g., earnings by occupation, race, age, sex). Difficult to obtain access to tapes.

4. Data on inter-relationships between individuals and groups in order to be able to predict outcomes of organizational changes.

5. Longitudinal data on enlisted groups by rating to analyze lifetime distributions under various policy conditions.

Figure III-6: Answers to Item 3: Describe briefly data you would like but have not been able to obtain.

previous entries for the same item are purged--except as some efforts have been made to save tapes--longitudinal studies have been difficult or impossible to conduct."

To summarize the questionnaire results as they apply to the question of demand for research data we can say that;

1) the type of data demanded is highly variable can be roughly categorized into three types shown on page 14 ,

2) much data is generated by individual research groups by questionnaire responses,

3) there is a clear need for longitudinal data,

4) both individual and group data is required.

The questionnaire proved useful in clarifying the scope of the data requirements of researchers, but was of little use in determining the frequency with which various data types might be accessed were it centrally available. To quote again from the letter of Dr. Glickman, "...I would encourage your efforts to refine estimates of the frequency with which items are likely to be used. The temptation to burden the system with "nice to know, just in case" data is great, and as you know, can put an impossible load on a system that tries to be correct and current..."

Before any attempt can be made to "refine" estimates of demand let us try to determine order of magnitude data request rates which a research data base might experience. Most manpower and personnel research on Navy and Marine Corps problems can be expected to be carried out at CNA, the Naval Personnel Research and Development Center in San Diego, and ONR contractors. Some work will be carried out by civilian contrators

with other branches of the Navy and Marine Corps and in other related areas such as medical personnel research at BUMED. Also various DOD related organizations such as IDA might make requests. Let us say, as order of magnitude figures, that CNA might have as many as 20 manpower and personnel research projects under way at any one time, similarly 20 at the San Diego Research Center and 20 contractors with ONR. It is emphasized that these are merely order-of-magnitude figures; the true figures are closer to 20 than either 2 or 200. These would make a total of sixty projects, and if projects on related topics were added one can guess that about 100 research projects at any one time might be a reasonable figure for manpower and personnel research projects. The authors doubt that it will be more than this, and it could be as low as 30 to 50.

Each project will, of course, generate its own data requirements which will produce variability in the demand on the data base. But let us assume that on the average a research group makes a request for data once per month. This figure it is believed is on the high side. Twelve separate requests for data per year by one group would surely be an upper bound. Such a request rate would lead to a demand rate on the RDB of about 3 requests per day. It is believed that such a figure would be an upper bound and that the request rate, especially at the start, would be considerably smaller than this.

What would be the time between requests for a given data element? It would of course vary considerably with the data elements requested, but an average time between requests using the above figures and 100 types of data in a data base (a reasonable number in the early period of an RDB) would be about 1 month.

These figures, although rough and subject to many criticisms, do put the demand rate at an RDB in perspective. The demand rate will be very low in comparison to requests made at an ODB or SDB, but if reliable data is available in an RDB, one must trade off the possible advantages of insight and results obtained by research with the cost of keeping data which is accessed infrequently.

IV. COSTS AND RESOURCES FOR A RESEARCH DATA BASE.

In this section an attempt is made to determine how much data base capability one can obtain at various costs, and costs of various alternative systems are given. At one extreme an RDB would be essentially a storage center which would keep data for distribution on request. At the other extreme the RDB would have staff to manipulate data, write software for these manipulations, and in general perform extensive editing, merging and sorting procedures. The costs given here are not detailed cost estimates, which are outside the scope of this report. They should be taken as reasonable estimates which might be further defined by detailed cost estimating.

In order to get a feel for the resources which might be available for an RDB let us use Table 1. The total funding for all personnel related research is currently about $60 million. If ½% of this were set aside for and RDB to collect and make available the data collected by this funded research we would have available $300,000. One pertinent question then is, is it worth this amount to collate and organize the large volumes of data which obviously are generated by a $60 million research budget?

This question may not be meaningful without some knowledge of what this $300,000 will purchase in the way of data base capability. It is difficult to decide the capability it would buy, since it is assumed that an RDB, if it were established, would be attached to an already existing computer system, in which case the actual charging of costs to the RDB would be subject to many variations. However, what follows is an attempt to put the various costs into focus.

1. Hardware Costs.

Let us assume that if an RDB were established it would be at a location already having access to a Navy-owned computer, and since the actual CPU time of the RDB would not be extensive, that the main frame costs would be $ zero. What an RDB would need would be tape drives and control units. Current lease costs of tape drives are about $625/month per unit, and a control unit about $750/month. For tape copying, a minimum would be two drives with a control unit, for a total of $2000/month. If sorting, editing and merging were desired then 4 drives would be needed (maybe more depending on volume). Each control unit can take 8 drives. Thus a medium size system would cost about $3,250/month with one control unit. For extensive merging and editing two control units may be required and with 4 drives this would cost $4,000/month. In addition to tape drives and control units, disk space would also be required to give greater flexibility in manipulating the data. This would cost about $1600/month.

2. Software Costs.

It is presumed that the data base manipulations would be carried out using a language already available on the machine in question. Thus no additional leasing of software would be required.

3. Personnel.

The number of personnel would, of course, depend on the size of the RDB envisioned. The smallest operation would consist of one or two people knowledgeable in computer input/output operations who could copy tape, print or punch output, and do little or no sorting or merging which would require writing of software. The personnel would carefully monitor incoming data, prepare routine updates of data base content for circulation, and copy data as requested by RDB users.

The largest operation visualized would have about 6 people. This would allow for the writing of software to manipulate the data, to sort, merge, and produce statistical tables and other useful output for potential RDB users.

4. The Minimum RDB.

The minimum RDB proposed here would require the following:

2 people (GS9-11) with computer background

2 Tape Drive Units

1 Tape Control Unit,

for a total annual cost of about $60,000 exclusive of computer main frame and software costs, office space and storage space. Any amount less than this would not allow for reasonable data access, and such a system could give only minimal service.

5. The Maximum RDB.

One can easily imagine a gargantuan data base which tries to keep all forums of historical manpower and personnel data. In fact it is quite common for data bases to fall prey to such imaginations and grow to such a size that it is difficult to extract data. As is mentioned earlier in this report, such data bases are not considered in the RDB category, and here we attempt to put an upper bound on the data base envisioned. The maximum RDB proposed here would require the following:

6 people (including a head, two programmers, two clerks and secretary)

4 Tape Drive Units

2 Control Units

Disk space,

for a total annual cost of about $150,00 exclusive of computer main frame

and software costs, office space and storage space. These facilities should be capable of supplying all the researchers needs for data if we assume the RDB is not to be used as an SDB.

The above cost figures could, with some effort, be refined. But as we stated above, it is not the purpose of this report to present detailed proposals and cost studies. These costs should suffice to put the range of an RDB in perspective.

V. STRUCTURE OF A RESEARCH DATA BASE.

In this section some remarks are made concerning how an RDB should be structured. The first fact to recognize is that manpower and personnel data will be available to the data base from two very different sources, 1) the operating Navy*, 2) the results of surveys by research groups. The first type of data is historical accounting information which gives cross-sectional records of the personnel in the Navy at various time points. If all this information is kept (approximately 15 reels for every time point) the data base will soon be swamped with much inaccessible data. Consideration should be given at the start to the keeping of extracts from these tapes. One could keep all pieces of information on a small sample of personnel, or only a few pieces of information on all personnel. Both these alternatives could be pursued, but every effort should be made from the outset to reduce as much as possible the data which is kept. Even more important, one might even say crucial, is that every piece of data which is kept must be well documented. It would seem that such an obvious requirement hardly needs to be stated, but enormous quantities of data are currently in storage which will not, and often cannot, be accessed or used because no one is sure what they are. This aspect has been stressed elsewhere in this report and is repeated here for emphasis.

A principal reason for the formation of an RDB is to bring together and make available from a central source the large amounts of data generated by research groups. Much of this data is in the form of answers to surveys, but it will include analyses of survey responses, the details of which are

---

*Included in the Operating Navy are such data sources as schools, medical facilities, etc.

not always found in research reports. Discussions with people at the Naval Personnel Research and Development Center indicate that they believe there is much duplication in the compilation of surveys because of the lack of a central source of data. They are undertaking their own study of a centralized data archive (see ref [4]).

Clearly the two types of data differ not only in source but in type. Consideration should be given to the formation of a "Statistical Operational Data Bank" and a "Behavioral Research Data Bank". In the latter case not only the survey results should be kept but also the survey questionnaires themselves, together with relevant information on their distribution. Much useful information should be available from existing data banks, such as that at the Survey Research Center in Michigan, on how such data should be stored.

Careful consideration should be given to the method of storing the data. It can be assumed that data requested from the RDB will be used in a wide variety of computer systems. The system should be flexible so that data can be copied in a variety of formats.

VI. IMPLEMENTATION AND DEVELOPMENT.

In the implementation of a new RDB it would be tempting to take all available manpower and personnel data currently in storage. This temptation should be resisted, and only that data which can be well documented and accessed should be used to start the data base. It is far better for the researcher to have access to small amounts of "clean" data, of trust worthy content and origin than to have available large masses of data whose origins and contents are suspect.

Although most of the points concerning the implementation and development of a data base are straight-forward and obvious a method of data base development deserves comments. With the large investment is manpower and personnel research it is obvious that the research undertaken generates large amounts of data. If an RDB is established one of the main justifications would be to recover and centralize this research-generated data. It should therefore be incumbent upon the researchers to make available in suitably documented form the data that they generate. This should be done in conjunction with the RDB staff to ensure compatibility of the data with the RDB system. The delivery to the RDB of relevant data could be made a part of the contract obligation of each research group. In this way the RDB becomes an important data source and would prove valuable in the avoidance of duplication of effort. It might also serve as a valuable source of information on what research is being done where. The author has found in the last few years in contacts with both researchers and operational personnel, that a recurring problem is the finding out of who is doing research in what areas and what has already been done. If most researchers were in contact with RDB staff it is possible that the existence of an RDB would help alleviate some of the confusion.

VII. LOCATION OF AN RDB.

Two locations present themselves for the location of an RDB, The Naval Personnel Research and Development Center in San Diego, and the Naval Postgraduate School in Monterey.

Both locations have the advantage of being away from the operational side of the Navy. The San Diego center already has large volumes of historical operational data and is currently undertaking a study of the formation of a survey research data bank. Maintenance of an RDB would seem to be a natural function for this center. A possible disadvantage might be that civilian institutions such as universities might encounter difficulties in communicating with the center, although the author has no direct evidence of any difficulties.

The Naval Postgraduate School has naturally closer ties with civilian universities. However, the operation and maintenance of an RDB is not an academic endeavor in itself and it would be difficult to locate it in the school structure. In comparison to the staff at the San Diego Center, the Postgraduate School faculty devotes relatively little effort to manpower and personnel research. At the time of writing the possibility of MARDAC moving part of its operations to Monterey is being considered. If this occurs, Navy RDB might have some interaction with it. However, the RDB envisioned would differ markedly in scope and size from MARDAC and it is not clear what interaction would be beneficial.

VIII. CONCLUSIONS.

The concept of a research data base (RDB) differs markedly from that of either an operating data base (ORB) or a staff data base (SDB). Although the demand for data from an RDB would be much less than that for an SDB, a case can be made for the existence of an RDB. Currently, valuable data generated by the considerable commitment of resources (currently about $59 million per year) in manpower and personnel research are lost, and this leads to duplication of effort and waste. A principle objective of an RDB would be to collect, store and make available the data generated by researchers to other researchers. In doing this it might prove a valuable focal point in determining the types of research being done by different groups.

Some rough cost figures are included in the report, and an RDB would operate in the range $60,000 - $150,000/year. Although these are crude figures it is felt that any RDB should operate within this range.

Two locations for an RDB are the Naval Personnel Research and Development Center in San Diego and the Naval Postgraduate School in Monterey. Both have certain advantages, but it would appear that the location in San Diego would be more appropriate.

REFERENCES

1. "Manpower Research and Management in Large Organizations", a report of the Task Force on Manpower Research, Defense Science Board, Office of Director of Defense Research and Engineering, Washington, D.C. 20301 (Dr. Eli Ginzberg, Chairman).

2. "MARDAC, Manpower Research and Data Analysis Center", Assistant Secretary of Defense (Manpower and Reserve Affairs), Washington, D. C. 20301.

3. "Bupers Futures Forum, 1975 - 1985", Bureau of Naval Personnel, Washington, D. C. 20370.

4. "Evaluation and Recommendations for Future Developments of the Proposed Attitudinal Information Data System" by M. H. Hansler and E. C. Bryant (Westat Inc.), obtain from the Naval Personnel Research and Development Center, San Diego 95152.

5. "Manpower and Personnel Data Bases in the Navy and Marine Corps", by K. T. Marshall, Report No. NPS55MT73021A, Naval Postgraduate School, Monterey, California 93940.

6. "Manpower Plans and Policy Division Cohort Data Base", Commandant, Marine Corps, Code MPI, Washington, D. C. 20380.

INITIAL DISTRIBUTION LIST

	No. copies
Defense Documentation Center Cameron Station, Building 5 5010 Duke Street Alexandria, Virginia 22314	12
Dean of Research Code 023 Naval Postgraduate School Monterey, California 93940	1
Library, Code 0212 Naval Postgraduate School Monterey, California 93940	2
Dr. Robert J. Lundegard, Code 430 Director, Mathematical and Information Sciences Division	1
LCDR Robert D. Matulka, Code 430C Research Program Officer	1
Dr. Thomas C. Varley, Code 434 Program Director Operations Research	1
Mr. Marvin Denicoff, Code 437 Program Director Information Systems	1
Dr. Glenn L. Bryan, Code 450 Director, Psychological Sciences Division	1
Dr. H. Wallace Sinaiko, Code 450 Research Study Director	1
Dr. John A. Nagay, Code 452 Director, Organizational Effectiveness Research Programs	1
Dr. Bert T. King, Code 452 Associate Director	1
Dr. Martin A. Tolcott, Code 455 Director Engineering Psychology Programs	1
Dr. Marshall J. Farr, Code 458 Director, Personnel and Training Research Programs	1
Mr. Robert J. Miller, Code 462 Director, Naval Analysis Programs	1
Mr. J. Randolph Simpson, Code 462 Supervisory Operations Research Analyst Office of Naval Research Arlington, Virginia 22217	1

Assistant Chief for Management Information 1
Bureau of Naval Personnel (Pers N)
Washington, D. C. 20370

Naval Personnel and Training Laboratory 1
San Diego, California 92152
Attention: Mr. J. Silverman, Program Director

Commanding Officer 1
Naval Personnel Research and Development Laboratory
Washington, D. C. 20390
Attention: Mr. A. L. Blanks

Supply and Fiscal Department 1
Code 514.32
NAVSTA
Washington, D. C. 20390

Assistant Deputy Chief for Manpower, OP102 1
Office of Chief of Naval Operations
Washington, D. C. 20390

Chief of Naval Training 1
Naval Training Command Data Services Center
Pensacola, Florida 32508

Rosalie K. Ambler 1
Head, Personnel Research Division
Naval Aerospace Medical Institute
Pensacola, Florida 32508

Manpower Management Information Systems Branch 1
Code A01M
Headquarters Marine Corps
Washington, D. C. 20380

Director, Manpower Requirements Division 1
Bureau of Medicine and Surgery
Washington, D. C. 20390

LCDR W. F. Ford, Director 1
Recruiting Data System
Navy Recruiting Command
Washington, D. C. 20370

Department of the Army 1
D.A.P.E.
Washington, D. C. 20310

L. P. Granning, Chief 1
ADP Division
U.S.A. Recruiting Command
Hampton, Virginia 23369

Mr. D. G. Redmond, Code 311 1
Maintenance Support Office
Mechanicsburg, Pennsulvania 17055

Chief of Naval Personnel 1
Pers N
Department of the Navy
Washington, D. C. 20370

Dr. Herschel Kanter, Director 1
Institute of Naval Studies
Center for Naval Analyses
1401 Wilson Boulevard
Arlington, Virginia 22209

Mr. H. M. Pitcher 1
Program Analysis Division
Institute for Defense Analyses
400 Army Navy Drive
Arlington, Virginia 22202

Dr. E. A. Fleishman 1
American Institutes for Research
8555 Sixteenth Street
Silver Spring, Maryland 20910

Mr. Philip G. Bernard 1
B-K Dynamics, Inc.
2351 Shady Grove Road
Rockville, Maryland 20850

Dr. Barry M. Feinberg 1
Bureau of Social Science Research, Inc.
1200 Seventeenth Street, N.W.
Washington, D. C. 20036

Professor Robert M. Oliver 1
University of California
Operations Research Center
Berkeley, California 94720

Dr. Richard S. Hatch 1
Decision Systems Associates, Inc.
11428 Rockville Pike
Rockville, Maryland 20852

Dr. Anita S. West
Denver Research Institute
University Park
Denver, Colorado 80210 — 1

Dr. Henry Solomon
George Washington University
Department of Economics
Washington, D. C. 20006 — 1

Mr. John P. Thomas
Hudson Institute
Quaker Ridge Road
Croton-on-Hudson, New York 10520 — 1

Mr. James N. Kelly
Management Analysis Center, Inc.
745 Concord Avenue
Cambridge, Massachusetts 02138 — 1

Dr. Lawrence Friedman
MATHEMATICA, Inc.
P. O. Box 2392
Princeton, New Jersey 08540 — 1

Dr. David G. Bowers
University of Michigan
Institute for Social Research
Ann Arbor, Michigan 48106 — 1

Dr. Jack R. Borsting
Department of Operations Research
and Administrative Sciences
Naval Postgraduate School
Monterey, California 93940 — 1

Mr. Michael W. Brown
Operations Research, Inc.
1400 Spring Street
Rockville, Maryland 20910 — 1

Dr. Marvin Dunnette
Personnel Decisions, Inc.
2515 Foshay Tower
Minneapolis, Minnesota 55402 — 1

Professor G. S. Watson
Princeton University
Department of Statistics
Princeton, New Jersey 08540 — 1

Mr. R. Bard Battelle 1
Stanford Research Institute
Naval Warfare Research Center
Menlo Park, California 94025

Dr. Gloria L. Grace 1
System Development Corporation
2500 Colorado Avenue
Santa Monica, California 90406

Dr. Leonard Carmichael 1
The Smithsonian Institution
Washington, D. C. 20560

Mr. Will E. Lassiter 1
Data Solutions Corporation
5272 River Road, Suite 100
Bethesda, Maryland 20016

Dr. Norman M. Abrahams 1
Naval Personnel and Training Research Laboratory
San Diego, California 92152

Dr. Herbert R. Northrup 1
Wharton School of Finance and Commerce
University of Pennsylvania
Philadelphia, Pennsylvania 19104

Professor Ezra S. Krendel 1
Department of Operations Research
University of Pennsylvania
Philadelphia, Pennsylvania 19104

Dr. Kneale T. Marshall 20
Department of Operations Research
and Administrative Sciences
Naval Postgraduate School
Monterey, California 93940

Assistant Chief for Research (Code 400)
Office of Naval Research
Arlington, Virginia 22217 — 1

Director of Research (Code 401)
Office of Naval Research
Arlington, Virginia 22217 — 1

Director (Code 460)
Naval Applications and Analysis Division
Office of Naval Research
Arlington, Virginia 22217 — 1

Research Psychologist
Office of Naval Research Branch Office
536 S. Clark St.
Chicago, Ill. 60605 — 1

Psychologist
Office of Naval Research Branch Office
495 Summer St.
Boston, Mass. 02210 — 1

Psychologist
Office of Naval Research Branch Office
1030 E. Green St.
Pasadena, CA 91106 — 1

Deputy Chief Scientist
Office of Naval Research Area Office
207 West 24th Street
New York, N. Y. 10011 — 1

Head of Manpower Training and
Reserve Group (Op-964D)
Room 4A538, The Pentagon
Washington, D. C. 20350 — 1

Assistant to the Assistant Deputy Chief
of Naval Operations (Manpower) (Op-01BZ2)
Room 4E473, The Pentagon
Washington, D. C. 20350 — 1

Deputy Director, Program Management Office
Naval Material Command (03PB)
Room 868, Crystal Plaza #6
2221 Jefferson Davis Highway
Arlington, Virginia 20360 — 1

Program Administrator, Personnel and Training Support 1
Naval Material Command (93424)
820 Crystal Plaza #6
2221 Jefferson Davis Highway
Arlington, Virginia 20360

Special Assistant for Enlisted Force Analysis 1
Naval Bureau of Personnel (Ax)
Room 2611, Arlington, Annex
Washington, D. C. 20370

Head, Project Volunteer Coordination Branch 1
Naval Bureau of Personnel (A25)
Room 2603, Arlington Annex
Washington, D. C. 20370

Technical Director, Personnel Research Division 1
Naval Bureau of Personnel (A3t)
Room 3077, Arlington Annex
Washington, D. C. 20370

Special Assistant to the Chief of Naval Personnel 1
Naval Bureau of Personnel (Oe)
Room 2403, Arlington Annex
Washington, D. C. 20370

Dr. M. Wiskoff
Naval Personnel Research and Development Center 1
San Diego, CA 92152

Dr. Richard C. Sorenson
Naval Personnel Research and Development Center 1
San Diego, CA 92152

Mr. Joe Silverman
Naval Personnel Research and Development Center 1
San Diego, CA 92152

Dr. Laurie Broedling
Naval Personnel Research and Development Center 1
San Diego, CA 92152

Dr. J. J. Collins (Op987F)
RDT&E Plans Division 1
Office of Chief of Naval Operations
Washington, D. C. 20350

Dr. Eli S. Flyer 1
Deputy for Research and Analysis
Assist. Sec. of Defense (M&RA)
Washington, D. C. 20301

Dr. R. A. Weitzman
Code 55
Naval Postgraduate School
Monterey, CA 93940 — 1

Dr. R. W. Butterworth
Code 55
Naval Postgraduate School
Monterey, CA 93940 — 1

Dr. P.A.W. Lewis
Code 55
Naval Postgraduate School
Monterey, CA 93940 — 1

Dr. D. P. Gaver
Code 55
Naval Postgraduate School
Monterey, CA 93940 — 1

U150175

DUDLEY KNOX LIBRARY - RESEARCH REPORTS
5 6853 01057765 3